Lecture 21 : Conditional Distributions

STAT205 *Lecturer: Jim Pitman* Scribe: *Richard Crump* <crump@econ.berkeley.edu>

21.1 Related Ideas

1. Regular conditional distribution for Y given X (or a sigma field $\sigma(X)$).
2. Markov kernel (transition function/operator).

21.1.1 Setup, Definitions and Terminology

Two measurable spaces: $(S, \mathcal{S})$ and $(T, \mathcal{T})$ (very often we have $S = T$ and $\mathcal{S} = \mathcal{T}$).

Definition 21.1 *A* Markov kernel *$P(s, A)$ from $(S, \mathcal{S})$ to $(T, \mathcal{T})$ is a collection of probability measures on $(T, \mathcal{T})$ indexed in an $\mathcal{S}$-measurable way by a parameter $s \in S$, satisfying:*

i. for each $s \in S$, $P(s, \cdot)$ is a probability measure on $(T, \mathcal{T})$; and

ii. for each $A \in \mathcal{T}$, $s \mapsto P(s, A)$ is measurable relative to $\mathcal{S}$.

Example 21.2 *From the theoretical statistics literature: $\mathbb{P}_\theta$ $\theta \in \Theta$. Where $\mathbb{P}_\theta$ are defined on the measurable space $(X, \mathcal{X})$.*

We now introduce the idea of a regular conditional distribution.

Definition 21.3 *For a pair of random variables (X, Y), $X : \Omega \mapsto S$ and $Y : \Omega \mapsto T$, defined on a background space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that a Markov kernel serves as a* regular conditional distribution *for Y given X if and only if the joint distribution of (X, Y) is given by the formula*

$$\mathbb{P}(X \in dx, Y \in dy) = \mathbb{P}(X \in dx) \cdot P(x, dy).$$

This means (by definition!) that

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \int_A \mathbb{P}(X \in dx) \cdot P(x, B) \qquad (21.1)\\
&= \int_S \mathbf{1}_{(x \in A)} \cdot \mathbb{P}(X \in dx) \cdot P(x, B)\\
&= \mathbb{E}\left[\mathbf{1}_{(X \in A)} \cdot P(X, B)\right].
\end{aligned}$$

Take $g(\cdot) \geq 0$ and product measurable: then,

$$\begin{aligned}
\mathbb{E}\left[g(X,Y)\right] &= \int_S \mathbb{P}(X \in dx)\left[\int_T P(x, dy) g(x,y)\right] \qquad (21.2)\\
&= \int_S \left[\int_T g(x,y) \cdot P(x, dy)\right] \mathbb{P}(X \in dx).
\end{aligned}$$

Theorem 21.4 (Generalized Fubini's Theorem) *(21.1) holds for all $A \in \mathcal{S},\ B \in \mathcal{T}$ if and only if (21.2) holds for all $g \geq 0$ product measurable (and hence for signed g, provided result is finite with $g \mapsto |g|$).*

21.1.2 Examples

Example 21.5 *If X, Y have density $f(x,y)$ relative to dx, dy for some reference measures dx on $(S, \mathcal{S})$ and dy on $(T, \mathcal{T})$, then*

$$P(x, dy) = \frac{f(x,y)}{f_X(x)} dy$$

In this case it is easy to flip and get a conditional distribution of X given Y, that is

$$\hat{P}(y, dx) = \frac{f(x,y)}{f_Y(y)} dx$$

by Bayes' Rule. But in a general setting (setting of Generalized Fubini's Theorem cannot do this), there is no Bayes' formula without reference measures. But, there is a general theory which says that provided $(S, \mathcal{S})$ is a nice, measurable space ([1], chapter 4, (1.6)) then there does exist a Markov kernel $\hat{P}(y, dx)$ so that

$$\mathbb{P}(X \in dx, Y \in dy) = \mathbb{P}(Y \in dy) \cdot \hat{P}(y, dx)$$

in the same sense as before.

Example 21.6 *A point* (X, Y) *is picked proportional to length measure on the perimeter of an equilateral triangle with vertices* $(0,0)$, $(1,0)$ *and* $(\frac{1}{2}, \frac{\sqrt{3}}{2})$. *Note that* (X, Y) *is easily defined on* $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}, \lambda)$.

Observe that $X \sim \mathcal{U}[0,1]$ *and* $Y \sim \frac{1}{3} \cdot \delta_0 + \frac{2}{3} \cdot \mathcal{U}[0, \frac{\sqrt{3}}{2}]$ *by inspection. What is the conditional distribution of* Y *given* $X = x$*?*

Note first that for $0 \leq x \leq \frac{1}{2}$ *the only possible values of* Y *consistent with* $X = x$ *are* $y = 0$ *or* $y = \sqrt{3}x$. *Thus*

$$\mathbb{P}(Y > 0 | 0 \leq x \leq \frac{1}{2}) = \frac{\mathbb{P}(Y > 0, 0 \leq x \leq \frac{1}{2})}{\mathbb{P}(0 \leq x \leq \frac{1}{2})} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6}} = \frac{2}{3}$$

More generally, choose $\varepsilon > 0$ *and* $0 < x < 1 - \varepsilon$. *Then,*

$$\mathbb{P}(Y > 0 | x < X < x + \varepsilon) = \frac{\mathbb{P}(Y > 0, x < X < x + \varepsilon)}{\mathbb{P}(x < X < x + \varepsilon)} = \frac{\frac{2}{3}\varepsilon}{\varepsilon} = \frac{2}{3}$$

and

$$\mathbb{P}(Y = 0 | x < X < x + \varepsilon) = \frac{1}{3}$$

So for $X = x$ *where* $0 \leq x \leq \frac{1}{2}$ *the conditional distribution is*

$$(Y | X = x) = \begin{cases} 0 & \text{with probability } \frac{1}{3} \\ \sqrt{3}x & \text{with probability } \frac{2}{3} \end{cases}$$

A similar argument follows for $\frac{1}{2} \leq x \leq 1$ *with a different line.*

Proof: You may check the details. Just check $\mathbb{P}(X \in A, Y \in B)$ is given by integration of this regular conditional distribution. Compute the lengths of intervals. ■

Question: How does this relate to conditional expectation?

Answer: Suppose $P(x, \cdot)$ gives a regular conditional distribution for Y given $X = x$ as defined before, i.e., $\mathbb{P}(X \in x, Y \in dy) = \mathbb{P}(X \in x) \cdot P(x, dy)$ then we can compute

$$\mathbb{E}[h(Y) | X = x] = \int_S P(x, dy) \cdot h(y) \equiv \psi(x)$$

so that

$$\mathbb{E}[h(Y) | X] = \psi(X)$$

and

$$\mathbb{E}[h(Y)] = \mathbb{E}[\psi(X)]$$

References

[1] Richard Durrett. *Probability: theory and examples, 3rd edition.* Thomson Brooks/Cole, 2005.